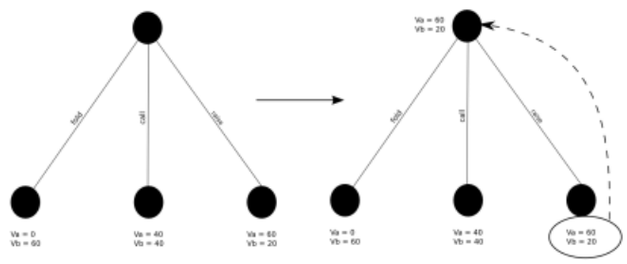
Reinforcement Learning in Poker

Our idea is to build a bot which have beliefs about his opponent hand. This beliefs are updated along the game according to differents methods. Each method give us a basic strategy and we use an UCB-like algorithm to choose the current strategy to use in game.

1. Game Tree:

First we generate a Min-Max tree using 4 different hands. The players real hands A0 and B0 and players beliefs on his opponent hand A1 and B1.
Node evaluation is done following a min-max like update, with the use of chance nodes for the community cards. Each nodes got 2 values (Va and Vb) : the amount of chip won by each player against his belief on his opponent. The value update of action nodes is done as follow : As example, if active player is A :

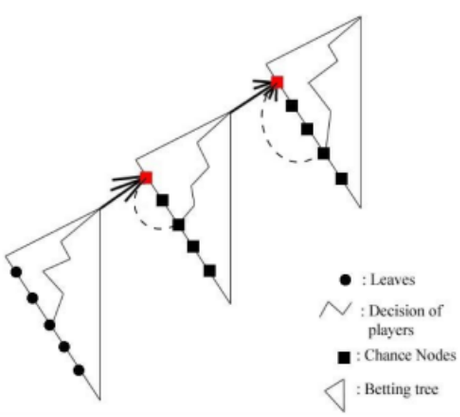


So we take the Max for the active player, and the value for the other player is the value corresponding to the choosen action.

We handle the different possibilities over B0 A1 B1 generating a forest of trees. B0 is uniformly drawn and A1, B1 comes from a beliefs vector which associate one hand with the probability of the player effectively own this hand conditionally to our model and the current position in the game tree.

2. Fixed Beliefs Assumption :

This forest of trees is too big to be handled so we use Monte Carlo methods and we decided to use an approximation : the fixed beliefs assumption. This means that current beliefs won't be updated during the exploration of the tree (but are still updated after opponent action).



Since the belief are fixed, we assume the trajectory followed in the betting trees are the same for each chance nodes reachable. The red nodes are fully computed and the values of black ones are inferred from the red.

Please note we don't do any hand clustering in our game reduction.

3. Belief Update :

After each opponent action the belief vector is updated according to a playing model. The probability for the hand H respectively to the past actions is :

$$P(H)_t = P(H)_{t-1} \cdot P(a|F,P,S)$$

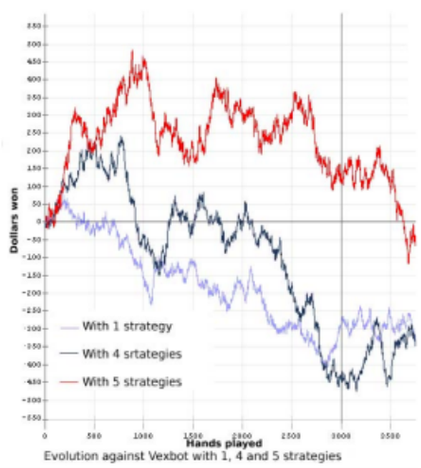
Were a is the action made by the opponent.
F the strength of the hand
H (brute winning probability)
P the amount of chips in the pot
S the stage of the game (flop, turn, or river)
P(a|F,P,S) is given by expert made playing models, we made combinaisons of these models to obtain several basis kinds of play such as agressive/passive/large/tight ones *et caetera*.

Our goal in this step is to allow sufficiently differents play styles to adapt to any opponent. Each time we identified a common biais in our basis strategies, we added a new strategie with the opposite biais. In our submited version for tournament we have 5 basics strategies.

4. Strategie Selection :

The choice of the current strategie if made by a UCB-like algorithm which allow our program to choose the correct strategie depending on the performance of each strategie. Moreover while we exploit a strategie we continue to explore the performances of the other ones.

We also aded to this part a change point detection to quickly adapt to variations of play of our opponent. This enables us to very quickly change our style of play when necessary.



Evolution against Vexbot with 1, 4 and 5 strategies

5. Results :

Results obtained against other programs in 3000 hands games

	Brennus	Vexbot	Sparbot	Always Call	Always Raise
Brennus		+0.05	+0.06	+1.01	+1.87
Vexbot	-0.05		+0.056	+1.04	+2.98
Sparbot	-0.06	-0.056		+0.47	+1.34
Always Call	-1.01	-1.04	-0.47		=0.00
Always Raise	-1.87	-2.98	-1.34	=0.00	

6. Main References :

[1] J.Y. Audibert, R. Munos and C. Szepesvari. Use of variance estimation in the multi armed bandit problem. NIPS 2006 Workshop on online trading of exploration and exploitation. Vancouver 2006.

[2] D. Billings, A. Davidson, T. Schauenberg, N. Burch, M. Bowling, R. Holte, J. Schaeffer, and D. Szafron. Game-tree search with adaptation in stochastic imperfect information games. In Computers and Games 4th International Conference, CG'04. LNCS 3846 pages 21-34.

Raphaël Maîtrepierre, Jérémie Mary, Rémi Munos
SEQUEL Team-Project, INRIA-FUTURS
Parc scientifique de la Haute-Borne
40 avenue Halley 59650 Villeneuve d'Ascq, FRANCE
r.maitrepierre@free.fr, {Jeremie.Mary, Remi.Munos}@inria.fr